Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context

Aki Ishii[a], Kazuyoshi Yata[b], Makoto Aoshima[b,1]

[a]*Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki, Japan*
[b]*Institute of Mathematics, University of Tsukuba, Ibaraki, Japan*

Abstract

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. We call such data HDLSS data. In this paper, we study asymptotic properties of the first principal component in the HDLSS context and apply them to equality tests of covariance matrices for high-dimensional data sets. We consider HDLSS asymptotic theories as the dimension grows for both the cases when the sample size is fixed and the sample size goes to infinity. We introduce an eigenvalue estimator by the noise-reduction methodology and provide asymptotic distributions of the largest eigenvalue in the HDLSS context. We construct a confidence interval of the first contribution ratio. We give asymptotic properties both for the first PC direction and PC score as well. We apply the findings to equality tests of two covariance matrices in the HDLSS context. We provide numerical results and discussions about the performances both on the estimates of the first PC and the equality tests of two covariance matrices.

Keywords: Contribution ratio, Equality test of covariance matrices, HDLSS, Noise-reduction methodology, PCA
2000 MSC: primary 34L20, secondary 62H25

Email address: aoshima@math.tsukuba.ac.jp (Makoto Aoshima)
[1]Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan; Fax: +81-298-53-6501

1. Introduction

One of the features of modern data is the data dimension d is high and the sample size n is relatively low. We call such data HDLSS data. In HDLSS situations such as $d/n \to \infty$, new theories and methodologies are required to develop for statistical inference based on the large sample theory. One of the approaches is to study geometric representations of HDLSS data and investigate the possibilities to make use of them in HDLSS statistical inference. Hall et al. (2005), Ahn et al. (2007), and Yata and Aoshima (2012) found several conspicuous geometric descriptions of HDLSS data when $d \to \infty$ while n is fixed. The HDLSS asymptotic studies usually assume either the normality as the population distribution or a ρ-mixing condition as the dependency of random variables in a sphered data matrix. See Jung and Marron (2009) and Jung et al. (2012). However, Yata and Aoshima (2009) developed an HDLSS asymptotic theory without assuming those assumptions and showed that the conventional principal component analysis (PCA) cannot give consistent estimation in the HDLSS context. In order to overcome this inconvenience, Yata and Aoshima (2012) provided the *noise-reduction (NR) methodology* that can successfully give consistent estimators of both the eigenvalues and eigenvectors together with the principal component (PC) scores. Furthermore, Yata and Aoshima (2010, 2013) created the *cross-data-matrix (CDM) methodology* that is a nonparametric method to ensure consistent estimation of those quantities. Given this background, Aoshima and Yata (2011, 2013) developed a variety of inference for HDLSS data such as given-bandwidth confidence region, two-sample test, test of equality of two covariance matrices, classification, variable selection, regression, pathway analysis and so on along with the sample size determination to ensure prespecified accuracy for each inference.

In this paper, suppose we have a $d \times n$ data matrix, $\boldsymbol{X}_{(d)} = [\boldsymbol{x}_{1(d)}, ..., \boldsymbol{x}_{n(d)}]$, where $\boldsymbol{x}_{j(d)} = (x_{1j(d)}, ..., x_{dj(d)})^T,\ j = 1, ..., n$, are independent and identically distributed (i.i.d.) as a d-dimensional distribution with a mean vector $\boldsymbol{\mu}_d$ and covariance matrix $\boldsymbol{\Sigma}_d$ $(\geq \boldsymbol{O})$. We assume $n \geq 3$. The eigen-decomposition of $\boldsymbol{\Sigma}_d$ is given by $\boldsymbol{\Sigma}_d = \boldsymbol{H}_d \boldsymbol{\Lambda}_d \boldsymbol{H}_d^T$, where $\boldsymbol{\Lambda}_d$ is a diagonal matrix of eigenvalues, $\lambda_{1(d)} \geq \cdots \geq \lambda_{d(d)} (\geq 0)$, and $\boldsymbol{H}_d = [\boldsymbol{h}_{1(d)}, ..., \boldsymbol{h}_{d(d)}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $\boldsymbol{X}_{(d)} - [\boldsymbol{\mu}_d, ..., \boldsymbol{\mu}_d] = \boldsymbol{H}_d \boldsymbol{\Lambda}_d^{1/2} \boldsymbol{Z}_{(d)}$. Then, $\boldsymbol{Z}_{(d)}$ is a $d \times n$ sphered data matrix from a distribution with the zero mean and the identity covariance matrix. Here, we write $\boldsymbol{Z}_{(d)} = [\boldsymbol{z}_{1(d)}, ..., \boldsymbol{z}_{d(d)}]^T$ and $\boldsymbol{z}_{j(d)} = (z_{j1(d)}, ..., z_{jn(d)})^T,\ j = 1, ..., d$. Note that $E(z_{ji(d)} z_{j'i(d)}) = 0$ $(j \neq j')$ and $\mathrm{Var}(\boldsymbol{z}_{j(d)}) = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. The i-th true PC

score of $\boldsymbol{x}_{j(d)}$ is given by $\boldsymbol{h}_{i(d)}^T(\boldsymbol{x}_{j(d)} - \boldsymbol{\mu}_d) = \lambda_{i(d)}^{1/2} z_{ij(d)}$ (hereafter called $s_{ij(d)}$). Note that $\mathrm{Var}(s_{ij(d)}) = \lambda_{i(d)}$ for all i, j. Hereafter, the subscript d will be omitted for the sake of simplicity when it does not cause any confusion. We assume that λ_1 has multiplicity one in the sense that $\liminf_{d\to\infty} \lambda_1/\lambda_2 > 1$. Also, we assume that $\limsup_{d\to\infty} E(z_{ij}^4) < \infty$ for all i, j and $P(\lim_{d\to\infty} ||\boldsymbol{z}_1|| \neq 0) = 1$. Note that if $\boldsymbol{X}$ is Gaussian, z_{ij}s are i.i.d. as the standard normal distribution, $N(0, 1)$. As necessary, we consider the following assumption for the normalized first PC scores, z_{1j} $(= s_{1j}/\lambda_1^{1/2})$, $j = 1, ..., n$:

(A-i) $\quad z_{1j},\ j = 1, ..., n$, are i.i.d. as $N(0, 1)$.

Note that $P(\lim_{d\to\infty} ||\boldsymbol{z}_1|| \neq 0) = 1$ under (A-i). Let us write the sample covariance matrix as $\boldsymbol{S} = (n-1)^{-1}(\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{X} - \overline{\boldsymbol{X}})^T = (n-1)^{-1}\sum_{j=1}^n (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T$, where $\overline{\boldsymbol{X}} = [\bar{\boldsymbol{x}}, ..., \bar{\boldsymbol{x}}]$ and $\bar{\boldsymbol{x}} = \sum_{j=1}^n \boldsymbol{x}_j/n$. Then, we define the $n \times n$ dual sample covariance matrix by $\boldsymbol{S}_D = (n-1)^{-1}(\boldsymbol{X} - \overline{\boldsymbol{X}})^T(\boldsymbol{X} - \overline{\boldsymbol{X}})$. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{n-1} \geq 0$ be the eigenvalues of $\boldsymbol{S}_D$. Let us write the eigendecomposition of $\boldsymbol{S}_D$ as $\boldsymbol{S}_D = \sum_{j=1}^{n-1} \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$, where $\hat{\boldsymbol{u}}_j = (\hat{u}_{j1}, ..., \hat{u}_{jn})^T$ denotes a unit eigenvector corresponding to $\hat{\lambda}_j$. Note that $\boldsymbol{S}$ and $\boldsymbol{S}_D$ share non-zero eigenvalues.

In this paper, we study asymptotic properties of the first principal component in the HDLSS context and apply them to equality tests of covariance matrices for high-dimensional data sets. We consider HDLSS asymptotic theories as $d \to \infty$ for both the cases when n is fixed and $n \to \infty$. In Section 2, we introduce an eigenvalue estimator by the NR methodology and provide asymptotic distributions of the largest eigenvalue in the HDLSS context. We construct a confidence interval of the first contribution ratio. In Section 3, we give asymptotic properties both for the first PC direction and PC score as well. In Section 4, we apply the findings to equality tests of two covariance matrices in the HDLSS context. Finally, in Section 5, we provide numerical results and discussions about the performances both on the estimates of the first PC and the equality tests of two covariance matrices.

2. Largest eigenvalue and its contribution rate

In this section, we give asymptotic distributions of the largest eigenvalue and construct a confidence interval of the first contribution rate.

2.1. Asymptotic distributions of the largest eigenvalue

Let $\delta_i = \mathrm{tr}(\boldsymbol{\Sigma}^2) - \sum_{s=1}^i \lambda_s^2 = \sum_{s=i+1}^d \lambda_s^2$ for $i = 1, ..., d-1$. We consider the following assumptions for the largest eigenvalue:

(A-ii) $\dfrac{\delta_1}{\lambda_1^2}=o(1)$ as $d\to\infty$ when n is fixed; $\dfrac{\delta_{i_*}}{\lambda_1^2}=o(1)$ as $d\to\infty$ for some fixed i_* $(<d)$ when $n\to\infty$.

(A-iii) $\dfrac{\sum_{r,s\ge 2}^{d}\lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\}}{n\lambda_1^2}=o(1)$ as $d\to\infty$ either when n is fixed or $n\to\infty$.

Note that (A-iii) holds when $\boldsymbol{X}$ is Gaussian and (A-ii) is met. Let $\boldsymbol{z}_{oj}=\boldsymbol{z}_j-(\bar{z}_j,...,\bar{z}_j)^T,\ j=1,...,p$, where $\bar{z}_j=n^{-1}\sum_{k=1}^{n}z_{jk}$. Let $\kappa=\mathrm{tr}(\boldsymbol{\Sigma})-\lambda_1=\sum_{s=2}^{d}\lambda_s$. Then, we have the following result.

Proposition 2.1. *Under (A-ii) and (A-iii), it holds that*

$$\frac{\hat{\lambda}_1}{\lambda_1}-||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2-\frac{\kappa}{\lambda_1(n-1)}=o_p(1)$$

as $d\to\infty$ either when n is fixed or $n\to\infty$.

Remark 2.1. Jung et al. (2012) gave a result similar to Proposition 2.1 when $\boldsymbol{X}$ is Gaussian, $\boldsymbol{\mu}=\boldsymbol{0}$ and n is fixed.

It holds that $E(||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2)=1$ and $||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2=1+o_p(1)$ as $n\to\infty$. If $\kappa/(n\lambda_1)=o(1)$ as $d\to\infty$ and $n\to\infty$, $\hat{\lambda}_1$ is a consistent estimator of λ_1. When n is fixed, the condition '$\kappa/\lambda_1=o(1)$' is equivalent to '$\lambda_1/\mathrm{tr}(\boldsymbol{\Sigma})=1+o(1)$' in which the contribution ratio of the first principal component is asymptotically 1. In that sense, '$\kappa/\lambda_1=o(1)$' is quite strict condition in real high-dimensional data analyses. Hereafter, we assume $\liminf_{d\to\infty}\kappa/\lambda_1>0$.

Yata and Aoshima (2012) proposed a method for eigenvalue estimation called the *noise-reduction (NR) methodology* that was brought by a geometric representation of $\boldsymbol{S}_D$. If one applies the NR methodology to the present case, λ_is are estimated by

$$\tilde{\lambda}_i=\hat{\lambda}_i-\frac{\mathrm{tr}(\boldsymbol{S}_D)-\sum_{j=1}^{i}\hat{\lambda}_j}{n-1-i}\quad(i=1,...,n-2). \tag{2.1}$$

Note that $\tilde{\lambda}_i\ge 0$ w.p.1 for $i=1,...,n-2$. Also, note that the second term in (2.1) with $i=1$ is an estimator of $\kappa/(n-1)$. See Lemma 2.1 in Section 2.2 for the details. Yata and Aoshima (2012, 2013) showed that $\tilde{\lambda}_i$ has several consistency properties when $d\to\infty$ and $n\to\infty$. On the other hand, Ishii et al. (2014) gave asymptotic properties of $\tilde{\lambda}_1$ when $d\to\infty$ while n is fixed. The following theorem summarizes their findings:

Theorem 2.1. *Under (A-ii) and (A-iii), it holds that as $d \to \infty$*

$$\frac{\tilde{\lambda}_1}{\lambda_1} = \begin{cases} ||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 + o_p(1) & \textit{when } n \textit{ is fixed}, \\ 1 + o_p(1) & \textit{when } n \to \infty. \end{cases}$$

Under (A-i) to (A-iii), it holds that as $d \to \infty$

$$(n-1)\frac{\tilde{\lambda}_1}{\lambda_1} \Rightarrow \chi^2_{n-1} \qquad \textit{when } n \textit{ is fixed},$$
$$\sqrt{\frac{n-1}{2}}\Big(\frac{\tilde{\lambda}_1}{\lambda_1} - 1\Big) \Rightarrow N(0,1) \quad \textit{when } n \to \infty.$$

Here, " $\Rightarrow$ " denotes the convergence in distribution and χ^2_{n-1} denotes a random variable distributed as χ^2 distribution with $n-1$ degrees of freedom.

2.2. Confidence interval of the first contribution ratio

We consider a confidence interval for the contribution ratio of the first principal component. Let a and b be constants satisfying $P(a \le \chi^2_{n-1} \le b) = 1 - \alpha$, where $\alpha \in (0,1)$. Then, from Theorem 2.1, under (A-i) to (A-iii), it holds that

$$\begin{aligned} &P\Big(\frac{\lambda_1}{\text{tr}(\boldsymbol{\Sigma})} \in \Big[\frac{(n-1)\tilde{\lambda}_1}{b\kappa + (n-1)\tilde{\lambda}_1}, \frac{(n-1)\tilde{\lambda}_1}{a\kappa + (n-1)\tilde{\lambda}_1}\Big]\Big) \\ &= P\Big(a \le (n-1)\frac{\tilde{\lambda}_1}{\lambda_1} \le b\Big) = 1 - \alpha + o(1) \end{aligned} \tag{2.2}$$

as $d \to \infty$ when n is fixed. We need to estimate κ in (2.2). Here, we give a consistent estimator of κ by $\tilde{\kappa} = (n-1)(\text{tr}(\boldsymbol{S}_D) - \hat{\lambda}_1)/(n-2) = \text{tr}(\boldsymbol{S}_D) - \tilde{\lambda}_1$. Then, we have the following results.

Lemma 2.1. *Under (A-ii) and (A-iii), it holds that*

$$\frac{\tilde{\kappa}}{\kappa} = 1 + o_p(1) \quad \textit{and} \quad \frac{\tilde{\kappa}}{\lambda_1} = \frac{\kappa}{\lambda_1} + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Theorem 2.2. *Under (A-i) to (A-iii), it holds that*

$$P\Big(\frac{\lambda_1}{\textit{tr}(\boldsymbol{\Sigma})} \in \Big[\frac{(n-1)\tilde{\lambda}_1}{b\tilde{\kappa} + (n-1)\tilde{\lambda}_1}, \frac{(n-1)\tilde{\lambda}_1}{a\tilde{\kappa} + (n-1)\tilde{\lambda}_1}\Big]\Big) = 1 - \alpha + o(1) \tag{2.3}$$

as $d \to \infty$ when n is fixed.

Remark 2.2. From Theorem 2.1 and Lemma 2.1, under (A-ii) and (A-iii), it holds that $\text{tr}(\boldsymbol{S}_D)/\text{tr}(\boldsymbol{\Sigma}) = (\tilde{\kappa} + \tilde{\lambda}_1)/\text{tr}(\boldsymbol{\Sigma}) = 1 + o_p(1)$ as $d \to \infty$ and $n \to \infty$. We have that

$$\frac{\tilde{\lambda}_1}{\text{tr}(\boldsymbol{S}_D)} = \frac{\lambda_1}{\text{tr}(\boldsymbol{\Sigma})}\{1 + o_p(1)\}.$$

Remark 2.3. The constants $(a,\ b)$ should be chosen for (2.3) to have the minimum length. If $\lambda_1/\kappa = o(1)$, the length of the confidence interval becomes close to $\{(n-1)\tilde{\lambda}_1/\tilde{\kappa}\}(1/a - 1/b)$ under (A-ii) and (A-iii) when $d \to \infty$ and n is fixed. Thus, we recommend to choose constants $(a,\ b)$ such that

$$\underset{a,b}{\text{argmin}}(1/a - 1/b) \quad \text{subject to} \;\; G_{n-1}(b) - G_{n-1}(a) = 1 - \alpha,$$

where $G_{n-1}(\cdot)$ denotes the c.d.f. of χ^2_{n-1}.

Let us construct a confidence interval for the contribution ratio of the first principal component. We used gene expression data by Armstrong et al. (2002) in which the data set consists of $12582\ (= d)$ genes. The data set has three leukemia subtypes: 24 samples from acute lymphoblastic leukemia (ALL), 20 samples from mixed-lineage leukemia (MLL), and 28 samples from acute myeloid leukemia (AML). We standardized each sample so as to have the unit variance. Then, it holds $\text{tr}(\boldsymbol{S})\ (= \text{tr}(\boldsymbol{S}_D)) = d$, so that $\tilde{\lambda}_1 + \tilde{\kappa} = d$. From Theorem 2.2, we constructed a 95% confidence interval of the first contribution rate for each data set by choosing $(a,\ b)$ as in Remark 2.3. The results are summarized in Table 1.

Table 1. The 95% confidence interval (CI) of the first contribution ratio, together with $\tilde{\lambda}_1$ and $\tilde{\kappa}$, for Armstrong et al. (2002)'s data sets having $d = 12582$.

	CI	$\tilde{\lambda}_1$	$\tilde{\kappa}$
ALL $(n = 24)$	$[0.0557, 0.1663]$	1256	11326
MLL $(n = 20)$	$[0.1201, 0.3458]$	2717	9865
AML $(n = 28)$	$[0.0706, 0.1884]$	1501	11081

3. First PC direction and PC score

In this section, we give asymptotic properties of the first PC direction and PC score in the HDLSS context.

3.1. *Asymptotic properties of the first PC direction*

Let $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{h}}_1, ..., \hat{\boldsymbol{h}}_d]$, where $\hat{\boldsymbol{H}}$ is a $d \times d$ orthogonal matrix of the sample eigenvectors such that $\hat{\boldsymbol{H}}^T \boldsymbol{S} \hat{\boldsymbol{H}} = \hat{\boldsymbol{\Lambda}}$ having $\hat{\boldsymbol{\Lambda}} = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_d)$. We assume $\boldsymbol{h}_i^T \hat{\boldsymbol{h}}_i \geq 0$ w.p.1 for all i without loss of generality. Note that $\hat{\boldsymbol{h}}_i$ can be calculated by $\hat{\boldsymbol{h}}_i = \{(n-1)\hat{\lambda}_i\}^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_i$. First, we have the following result.

Lemma 3.1. *Under (A-ii) and (A-iii), it holds that*

$$\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1 - \Big(1 + \frac{\kappa}{\lambda_1 ||\boldsymbol{z}_{o1}||^2}\Big)^{-1/2} = o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

If $\kappa/(n\lambda_1) = o(1)$ as $d \to \infty$ and $n \to \infty$, $\hat{\boldsymbol{h}}_1$ is a consistent estimator of $\boldsymbol{h}_1$ in the sense that $\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1 = 1 + o_p(1)$. When n is fixed, $\hat{\boldsymbol{h}}_1$ is not a consistent estimator because $\lim_{d\to\infty} \kappa/\lambda_1 > 0$. In order to overcome this inconvenience, we consider applying the NR methodology to the PC direction vector. Let $\tilde{\boldsymbol{h}}_i = \{(n-1)\tilde{\lambda}_i\}^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_i$. From Lemma 3.1, we have the following result.

Theorem 3.1. *Under (A-ii) and (A-iii), it holds that*

$$\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1 = 1 + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Note that $||\tilde{\boldsymbol{h}}_1||^2 = \hat{\lambda}_1/\tilde{\lambda}_1 \geq 1$ w.p.1. We emphasize that $\tilde{\boldsymbol{h}}_1$ is a consistent estimator of $\boldsymbol{h}_1$ in the sense of the inner product even when n is fixed though $\tilde{\boldsymbol{h}}_1$ is not a unit vector. We give an application of $\tilde{\boldsymbol{h}}_1$ in Section 4.

3.2. *Asymptotic properties of the first PC score*

Let $z_{oij} = z_{ij} - \bar{z}_i$ for all i, j. First, we have the following result.

Lemma 3.2. *Under (A-ii) and (A-iii), it holds that*

$$\hat{u}_{1j} = z_{o1j}/||\boldsymbol{z}_{o1}|| + o_p(1) \quad \textit{for } j = 1, ..., n$$

as $d \to \infty$ when n is fixed.

Remark 3.1. By using Lemma 3.2 and the test of normality such as Jarque-Bera test, one can check whether (A-i) holds or not.

By applying the NR methodology to the first PC score, we obtain an estimate by $\tilde{s}_{1j} = \sqrt{(n-1)\tilde{\lambda}_1}\hat{u}_{1j},\ j = 1, ..., n$. A sample mean squared error of the first PC score is given by $\text{MSE}(\tilde{s}_1) = n^{-1}\sum_{j=1}^{n}(\tilde{s}_{1j} - s_{1j})^2$. Then, from Theorem 2.1 and Lemma 3.2, we have the following result.

Theorem 3.2. *Under (A-ii) and (A-iii), it holds that*

$$\frac{1}{\sqrt{\lambda_1}}(\tilde{s}_{1j} - s_{1j}) = -\bar{z}_1 + o_p(1) \quad \textit{for } j = 1, ..., n$$

as $d \to \infty$ when n is fixed. Under (A-i) to (A-iii), it holds that

$$\sqrt{\frac{n}{\lambda_1}}(\tilde{s}_{1j} - s_{1j}) \Rightarrow N(0, 1) \quad \textit{for } j = 1, ..., n; \quad \textit{and} \quad n\frac{MSE(\tilde{s}_1)}{\lambda_1} \Rightarrow \chi_1^2$$

as $d \to \infty$ when n is fixed.

Remark 3.2. The conventional estimator of the first PC score is given by $\hat{s}_{1j} = \sqrt{(n-1)\hat{\lambda}_1}\hat{u}_{1j},\ j = 1, ..., n$. From Theorems 8.1 and 8.2 in Yata and Aoshima (2013), under (A-ii) and (A-iii), it holds that as $d \to \infty$ and $n \to \infty$

$$\frac{\text{MSE}(\hat{s}_1)}{\lambda_1} = o_p(1) \ \text{ if } \kappa/(n\lambda_1) = o(1), \quad \text{and} \quad \frac{\text{MSE}(\tilde{s}_1)}{\lambda_1} = o_p(1).$$

4. Equality tests of two covariance matrices

In this section, we consider the test of equality of two covariance matrices in the HDLSS context. Even though there are a variety of tests to deal with covariance matrices when $d \to \infty$ and $n \to \infty$, there seem to be no tests available in the HDLSS context such as $d \to \infty$ while n is fixed. Suppose we have two independent $d \times n_i$ data matrices, $\boldsymbol{X}_i = [\boldsymbol{x}_{1(i)}, ..., \boldsymbol{x}_{n_i(i)}],\ i = 1, 2$, where $\boldsymbol{x}_{j(i)},\ j = 1, ..., n_i$, are i.i.d. as a d-dimensional distribution, π_i, having a mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i\ (\geq \boldsymbol{O})$. We assume $n_i \geq 3,\ i = 1, 2$. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ is given by $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i\boldsymbol{\Lambda}_i\boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{1(i)}, ..., \lambda_{d(i)})$ having $\lambda_{1(i)} \geq \cdots \geq \lambda_{d(i)} (\geq 0)$ and $\boldsymbol{H}_i = [\boldsymbol{h}_{1(i)}, ..., \boldsymbol{h}_{d(i)}]$ is an orthogonal matrix of the corresponding eigenvectors.

4.1. Equality test using the largest eigenvalues

We consider the following test for the largest eigenvalues:

$$H_0 : \lambda_{1(1)} = \lambda_{1(2)} \quad \text{vs.} \quad H_a : \lambda_{1(1)} \neq \lambda_{1(2)} \ \ (\text{or} \ \ H_b : \lambda_{1(1)} < \lambda_{1(2)}). \tag{4.1}$$

Let $\tilde{\lambda}_{1(i)}$ be the estimate of $\lambda_{1(i)}$ by the NR methodology as in (2.1) for π_i. Let $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. From Theorem 2.1, we have the following result.

Corollary 4.1. *Under (A-i) to (A-iii) for each π_i, it holds that*

$$\frac{\tilde{\lambda}_{1(1)}/\lambda_{1(1)}}{\tilde{\lambda}_{1(2)}/\lambda_{1(2)}} \Rightarrow F_{\nu_1,\nu_2}$$

as $d \to \infty$ when n_is are fixed, where F_{ν_1,ν_2} denotes a random variable distributed as F distribution with degrees of freedom, ν_1 and ν_2.

Let $F_1 = \tilde{\lambda}_{1(1)}/\tilde{\lambda}_{1(2)}$. From Corollary 4.1, we test (4.1) for given $\alpha \in (0, 1/2)$ by

$$\text{accepting } H_a \Longleftrightarrow F_1 \notin [\{F_{\nu_2,\nu_1}(\alpha/2)\}^{-1}, F_{\nu_1,\nu_2}(\alpha/2)] \tag{4.2}$$
$$\text{or} \quad \text{accepting } H_b \Longleftrightarrow F_1 < \{F_{\nu_2,\nu_1}(\alpha)\}^{-1}, \tag{4.3}$$

where $F_{\nu_1,\nu_2}(\alpha)$ denotes the upper $\alpha\%$ point of F distribution with degrees of freedom, ν_1 and ν_2. Then, under (A-i) to (A-iii) for each π_i, it holds that

$$\text{size} = \alpha + o(1)$$

as $d \to \infty$ when n_is are fixed.

Now, we check the performance of the test by (4.2) or (4.3). We also consider a test by the conventional estimator, $\hat{\lambda}_{1(i)}$. Let $\kappa_i = \text{tr}(\boldsymbol{\Sigma}_i) - \lambda_{1(i)} = \sum_{s=2}^{d} \lambda_{s(i)}$ for $i = 1, 2$. From Proposition 2.1, if $\kappa_i/\lambda_{1(i)} = o(1)$, $i = 1, 2$, under (A-i) to (A-iii) for each π_i it holds that

$$\frac{\hat{\lambda}_{1(1)}/\lambda_{1(1)}}{\hat{\lambda}_{1(2)}/\lambda_{1(2)}} \Rightarrow F_{\nu_1,\nu_2}$$

as $d \to \infty$ when n_is are fixed. As mentioned in Section 2, the condition '$\kappa_i/\lambda_{1(i)} = o(1)$ for $i = 1, 2$' is quite strict in real high-dimensional data analyses. Hereafter, we assume $\liminf_{d\to\infty} \kappa_i/\lambda_{1(i)} > 0$ for $i = 1, 2$. We analyzed the same gene expression data as in Table 1. We set $\alpha = 0.05$. We considered two cases: (I)

π_1 : ALL ($n_1 = 24$) and π_2 : MLL ($n_2 = 20$), and (II) π_1 : AML ($n_1 = 28$) and π_2 : MLL ($n_2 = 20$). As for $F_1' = \hat{\lambda}_{1(1)}/\hat{\lambda}_{1(2)}$, we considered (4.2) and (4.3) by replacing F_1 with F_1'. The results are summarized in Table 2. We observed from Table 2 that only H_b for (I) was accepted by F_1, namely, only F_1 for (I) rejected H_0 vs. H_b. One should note that the condition '$\kappa_i/\lambda_{1(i)} = o(1)$ for $i = 1, 2$' does not hold both for (I) and (II) as observed in Table 1.

Table 2. Tests of $H_0 : \lambda_{1(1)} = \lambda_{1(2)}$ vs. $H_a : \lambda_{1(1)} \neq \lambda_{1(2)}$ or $H_b : \lambda_{1(1)} < \lambda_{1(2)}$ with size 0.05 for Armstrong et al. (2002)'s data sets having $d = 12582$.

	H_a by F_1	H_a by F_1'	H_b by F_1	H_b by F_1'
(I) π_1: ALL, π_2: MLL	Reject	Reject	Accept	Reject
(II) π_1: AML, π_2: MLL	Reject	Reject	Reject	Reject

4.2. Equality test using the largest eigenvalues and their PC directions

We consider the following test using the largest eigenvalues and their PC directions:

$$H_0 : (\lambda_{1(1)}, \boldsymbol{h}_{1(1)}) = (\lambda_{1(2)}, \boldsymbol{h}_{1(2)}) \quad \text{vs.} \quad H_a : (\lambda_{1(1)}, \boldsymbol{h}_{1(1)}) \neq (\lambda_{1(2)}, \boldsymbol{h}_{1(2)}). \tag{4.4}$$

Let $\tilde{\boldsymbol{h}}_{1(i)}$ be the estimator of the first PC direction for π_i by the NR methodology given in Section 3.1. We assume $\boldsymbol{h}_{1(i)}^T \tilde{\boldsymbol{h}}_{1(i)} \geq 0$ w.p.1 for $i = 1, 2$, without loss of generality. Here, we have the following result.

Lemma 4.1. *Under (A-ii) and (A-iii) for each π_i, it holds that*

$$\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)} = \boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)} + o_p(1)$$

as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$.

Let $\tilde{h} = |\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)}|/2 + |\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)}|^{-1}/2$. Note that $\tilde{h} \geq 1$. Then, from Lemma 4.1, we give a test statistic for (4.4) as follows:

$$F_2 = \frac{\tilde{\lambda}_{1(1)}}{\tilde{\lambda}_{1(2)}} \tilde{h}_*,$$

where

$$\tilde{h}_* = \begin{cases} \tilde{h} & \text{if } \tilde{\lambda}_{1(1)} \geq \tilde{\lambda}_{1(2)}, \\ \tilde{h}^{-1} & \text{otherwise}. \end{cases}$$

From Lemma 4.1, we have the following result.

Theorem 4.1. *Under (A-i) to (A-iii) for each π_i, it holds that*

$$F_2 \Rightarrow F_{\nu_1,\nu_2} \text{ under } H_0$$

as $d \to \infty$ when n_is are fixed.

From Theorem 4.1, we consider testing (4.4) by (4.2) with F_2 instead of F_1. Then, the size becomes close to α as d increases. For the same gene expression data sets as in Section 4.1, we tested (4.4) with $\alpha = 0.05$ for the cases of (I) and (II). We observed that only H_a for (II) was accepted by F_2, namely, only F_2 for (II) rejected H_0 vs. H_a in (4.4).

4.3. Equality test of the covariance matrices

We consider the following test for the covariance matrices:

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \quad \text{vs.} \quad H_a : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2. \tag{4.5}$$

When $d \to \infty$ and n_is are fixed, one cannot estimate $\lambda_{j(i)}$s and $\boldsymbol{h}_{j(i)}$s for $j = 2, ..., d$. Instead, we consider estimating κ_is. Let $\boldsymbol{S}_{D(i)}$ be the dual sample covariance matrix for π_i. We estimate κ_i by $\tilde{\kappa}_i = \text{tr}(\boldsymbol{S}_{D(i)}) - \tilde{\lambda}_{1(i)}$ for $i = 1, 2$. From Lemma 2.1, under (A-ii) and (A-iii) for each π_i, $\tilde{\kappa}_i$s are consistent estimators of κ_is in the sense that $\tilde{\kappa}_i/\kappa_i = 1 + o_p(1)$ as $d \to \infty$ when n_is are fixed. Let $\tilde{\gamma} = \max\{\tilde{\kappa}_1/\tilde{\kappa}_2, \tilde{\kappa}_2/\tilde{\kappa}_1\}$. Now, we give a test statistic for (4.5) as follows:

$$F_3 = \frac{\tilde{\lambda}_{1(1)}}{\tilde{\lambda}_{1(2)}} \tilde{h}_* \tilde{\gamma}_*,$$

where

$$\tilde{\gamma}_* = \begin{cases} \tilde{\gamma} & \text{if } \tilde{\lambda}_{1(1)} \geq \tilde{\lambda}_{1(2)}, \\ \tilde{\gamma}^{-1} & \text{otherwise}. \end{cases}$$

Then, we have the following result.

Theorem 4.2. *Under (A-i) to (A-iii) for each π_i, it holds that*

$$F_3 \Rightarrow F_{\nu_1,\nu_2} \text{ under } H_0$$

as $d \to \infty$ when n_is are fixed.

From Theorem 4.2, we consider testing (4.5) by (4.2) with F_3 instead of F_1. Then, the size becomes close to α as d increases. For the same gene expression data sets as in Section 4.1, we tested (4.5) with $\alpha = 0.05$ for the cases of (I) and (II). We compared the performance of F_3 with two other test statistics: Q_2^2 and T_2^2 by Srivastava and Yanagihara (2010). The results are summarized in Table 3. We observed that H_a was accepted by F_3 both for (I) and (II), namely, F_3 rejected H_0 vs. H_a in (4.5) for both the cases. On the other hand, Q_2^2 and T_2^2 did not work for these data sets. It should be noted that Q_2^2 and T_2^2 require to meet the conditions that $0 < \lim_{d\to\infty} \text{tr}(\boldsymbol{\Sigma}^i)/d < \infty$ $(i = 1, ..., 4)$ and $d^{1/2}/n = o(1)$. As observed in Table 1, the conditions seem not to hold for these data sets with $d = 12582$ and $n \le 28$. Hence, there is no theoretical guarantee for the results by Q_2^2 and T_2^2.

Table 3. Tests of $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ vs. $H_a : \boldsymbol{\Sigma}_1 \ne \boldsymbol{\Sigma}_2$ with size 0.05 for Armstrong et al. (2002)'s data sets having $d = 12582$.

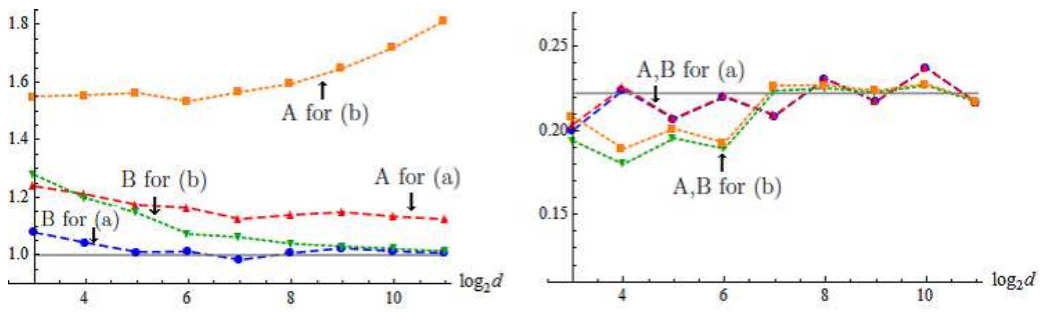
	H_a by F_3	H_a by Q_2^2	H_a by T_2^2
(I) π_1: ALL, π_2: MLL	Accept	Reject	Reject
(II) π_1: AML, π_2: MLL	Accept	Reject	Reject

5. Numerical results and discussions

5.1. Comparisons of the estimates on the first PC

In this section, we compared the performance of $\tilde{\lambda}_1$, $\tilde{\boldsymbol{h}}_1$ and $\tilde{s}_{1j}$ with their conventional counterparts by Monte Carlo simulations. We set $d = 2^k,\ k = 3, ..., 11$ and $n = 10$. We considered two cases for λ_is: (a) $\lambda_i = d^{1/i},\ i = 1, ..., d$ and (b) $\lambda_i = d^{3/(2+2i)},\ i = 1, ..., d$. Note that $\lambda_1 = d$ for (a) and $\lambda_1 = d^{3/4}$ for (b). Also, note that (A-ii) holds both for (a) and (b). Let $d_* = \lceil d^{1/2} \rceil$, where $\lceil x \rceil$ denotes the smallest integer $\ge x$. We considered a non-Gaussian distribution as follows: $(z_{1j}, ..., z_{d-d_*j})^T,\ j = 1, ..., n$, are i.i.d. as $N_{d-d_*}(\boldsymbol{0}, \boldsymbol{I}_{d-d_*})$ and $(z_{d-d_*+1j}, ..., z_{dj})^T,\ j = 1, ..., n$, are i.i.d. as the d_*-variate t-distribution, $t_{d_*}(\boldsymbol{0}, \boldsymbol{I}_{d_*}, 10)$ with mean zero, covariance matrix $\boldsymbol{I}_{d_*}$ and degrees of freedom 10, where $(z_{1j}, ..., z_{d-d_*j})^T$ and $(z_{d-d_*+1j}, ..., z_{dj})^T$ are independent for each j. Note that (A-i) and (A-iii) hold both for (a) and (b) from the fact that $\sum_{r,s\ge 2}^{d} \lambda_r \lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\} = 2\sum_{s=2}^{d-d_*} \lambda_s^2 + O(\sum_{r,s\ge d-d_*+1}^{d} \lambda_r \lambda_s) = o(\lambda_1^2)$.

The findings were obtained by averaging the outcomes from 2000 $(= R$, say) replications. Under a fixed scenario, suppose that the r-th replication ends with

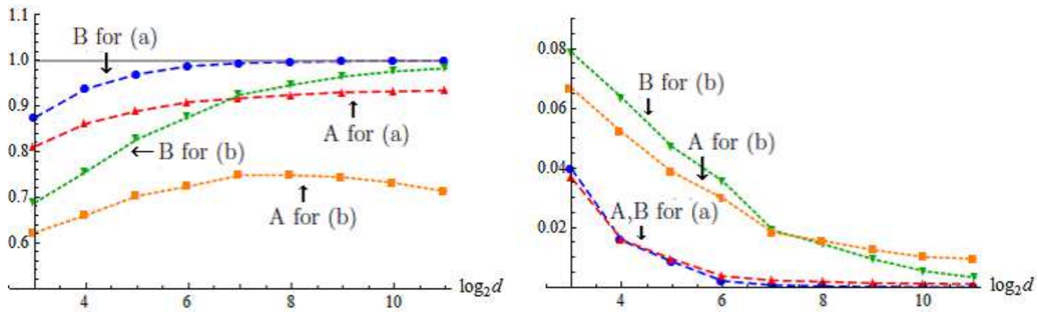


A: $\hat{\lambda}_1/\lambda_1$ and B: $\tilde{\lambda}_1/\lambda_1$ A: var$(\hat{\lambda}_1/\lambda_1)$ and B: var$(\tilde{\lambda}_1/\lambda_1)$

Figure 1. The values of A: $\hat{\lambda}_1/\lambda_1$ and B: $\tilde{\lambda}_1/\lambda_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The values of A: var$(\hat{\lambda}_1/\lambda_1)$ and B: var$(\tilde{\lambda}_1/\lambda_1)$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The asymptotic variance of $\tilde{\lambda}_1/\lambda_1$ was given by Var$\{\chi^2_{n-1}/(n-1)\} = 0.222$ and denoted by the solid line in the left panel.

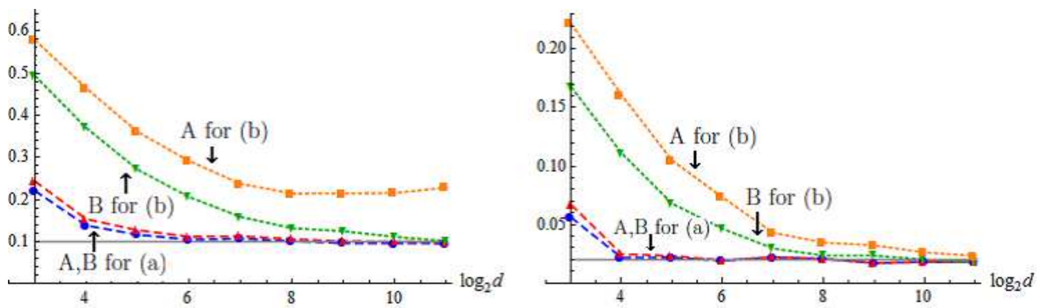
estimates, ($\hat{\lambda}_{1r}$, $\hat{\boldsymbol{h}}_{1r}$, MSE$(\hat{s}_1)_r$) and ($\tilde{\lambda}_{1r}$, $\tilde{\boldsymbol{h}}_{1r}$, MSE$(\tilde{s}_1)_r$) $(r = 1, ..., R)$. Let us simply write $\hat{\lambda}_1 = R^{-1}\sum_{r=1}^{R}\hat{\lambda}_{1r}$ and $\tilde{\lambda}_1 = R^{-1}\sum_{r=1}^{R}\tilde{\lambda}_{1r}$. We also considered the Monte Carlo variability by var$(\hat{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^{R}(\hat{\lambda}_{1r}-\hat{\lambda}_1)^2/\lambda_1^2$ and var$(\tilde{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^{R}(\tilde{\lambda}_{1r}-\tilde{\lambda}_1)^2/\lambda_1^2$. Figure 1 shows the behaviors of ($\hat{\lambda}_1/\lambda_1$, $\tilde{\lambda}_1/\lambda_1$) in the left panel and (var$(\hat{\lambda}_1/\lambda_1)$, var$(\tilde{\lambda}_1/\lambda_1)$) in the right panel for (a) and (b). We gave the asymptotic variance of $\tilde{\lambda}_1/\lambda_1$ by Var$\{\chi^2_{n-1}/(n-1)\} = 0.222$ from Theorem 2.1 and showed it by the solid line in the right panel. We observed that the sample mean and variance of $\tilde{\lambda}_1/\lambda_1$ become close to those asymptotic values as d increases.

Similarly, we plotted ($\hat{\boldsymbol{h}}_1^T\boldsymbol{h}_1$, $\tilde{\boldsymbol{h}}_1^T\boldsymbol{h}_1$) and (var$(\hat{\boldsymbol{h}}_1^T\boldsymbol{h}_1)$, var$(\tilde{\boldsymbol{h}}_1^T\boldsymbol{h}_1)$) in Figure 2 and (MSE$(\hat{s}_1)/\lambda_1$, MSE$(\tilde{s}_1)/\lambda_1$) and (var(MSE$(\hat{s}_1)/\lambda_1$), var(MSE$(\tilde{s}_1)/\lambda_1$)) in Figure 3. From Theorem 3.2, we gave the asymptotic mean of MSE$(\tilde{s}_1)/\lambda_1$ by $E(\chi^2_1/n) = 0.1$ and showed it by the solid line in the left panel of Figure 3. We also gave the asymptotic variance of MSE$(\tilde{s}_1)/\lambda_1$ by Var$(\chi^2_1/n) = 0.02$ in the right panel of Figure 3. Throughout, the estimators by the NR method gave good performances both for (a) and (b) when d is large. However, the conventional estimators gave poor performances especially for (b). This is probably because the bias of the conventional estimators, $\kappa/(n\lambda_1)$, is large for (b) compared to (a). See Proposition 2.1 for the details.



A: $\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ and B: $\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ A: var($\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$) and B: var($\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$)

Figure 2. The values of A: $\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ and B: $\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The values of A: var($\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$) and B: var($\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$) are denoted by the dashed lines for (a) and by the dotted lines for (b) in the right panel.



A: MSE$(\hat{s}_1)/\lambda_1$ and B: MSE$(\tilde{s}_1)/\lambda_1$ A: var(MSE$(\hat{s}_1)/\lambda_1$) and B: var(MSE$(\tilde{s}_1)/\lambda_1$)

Figure 3. The values of A: MSE$(\hat{s}_1)/\lambda_1$ and B: MSE$(\tilde{s}_1)/\lambda_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The values of A: var(MSE$(\hat{s}_1)/\lambda_1$) and B: var(MSE$(\tilde{s}_1)/\lambda_1$) are denoted by the dashed lines for (a) and by the dotted lines for (b) in the right panel. The asymptotic mean and variance of MSE$(\tilde{s}_1)/\lambda_1$ were given by $E(\chi_1^2/n) = 0.1$ and Var$(\chi_1^2/n) = 0.02$ and denoted by the solid lines in both the panels.

5.2. Equality tests of two covariance matrices

We used computer simulations to study the performance of the test procedures by F_1 for (4.1), F_2 for (4.4) and F_3 for (4.5). We set $\alpha = 0.05$. Independent pseudo-random normal observations were generated from $\pi_i : N_d(\mathbf{0}, \mathbf{\Sigma}_i)$, $i = 1, 2$. We set $(n_1, n_2) = (10, 20)$. We considered the cases: $d = 2^k,\ k = 3, ..., 11$, and

$$\mathbf{\Sigma}_i = \begin{pmatrix} \mathbf{\Sigma}_{i(1)} & \boldsymbol{O}_{2,d-2} \\ \boldsymbol{O}_{d-2,2} & \mathbf{\Sigma}_{i(2)} \end{pmatrix},\ i = 1, 2, \tag{5.1}$$

where $\boldsymbol{O}_{k,l}$ is the $k \times l$ zero matrix, $\mathbf{\Sigma}_{1(1)} = \text{diag}(d^{3/4}, d^{1/2})$ and $\mathbf{\Sigma}_{1(2)} = (0.3^{|s-t|})$. When considered the alternative hypotheses, we set

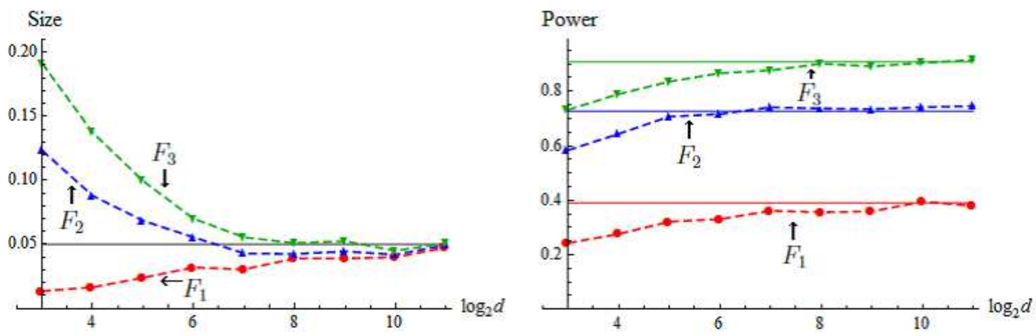
$$\mathbf{\Sigma}_{2(1)} = \begin{pmatrix} 1/3 & \sqrt{8}/3 \\ \sqrt{8}/3 & -1/3 \end{pmatrix} \text{diag}(3d^{3/4}, 1.5d^{1/2}) \begin{pmatrix} 1/3 & \sqrt{8}/3 \\ \sqrt{8}/3 & -1/3 \end{pmatrix} \tag{5.2}$$

and $\mathbf{\Sigma}_{2(2)} = 1.5(0.3^{|s-t|})$. Note that $\lambda_{1(2)}/\lambda_{1(1)} = 3$, $\kappa_2/\kappa_1 = 1.5$, $\boldsymbol{h}_{1(1)} = (1, 0, ..., 0)^T$ and $\boldsymbol{h}_{1(2)} = (1/3, \sqrt{8}/3, 0...., 0)^T$, so that $\boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)} = 1/3$. Also, note that (A-i) to (A-iii) hold for each π_i. Let $h = (|\boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)}| + 1/|\boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)}|)/2$ and $\gamma = \max\{\kappa_1/\kappa_2, \kappa_2/\kappa_1\}$. From Lemmas 2.1 and 4.1, it holds that $\tilde{h} = h + o_p(1)$ and $\tilde{\gamma} = \gamma + o_p(1)$. Thus, from Corollary 4.1, Theorems 4.1 and 4.2, we obtained the asymptotic powers of F_1, F_2 and F_3 with $(\tilde{h}_*, \tilde{\gamma}_*) = (h^{-1}, \gamma^{-1})$ as follows:

$$\begin{aligned}
&\text{Power}(F_1) = P\big\{(\lambda_{1(1)}/\lambda_{1(2)})f \notin [\{F_{\nu_2,\nu_1}(\alpha/2)\}^{-1}, F_{\nu_1,\nu_2}(\alpha/2)]\big\} = 0.39, \\
&\text{Power}(F_2) = P\big\{h^{-1}(\lambda_{1(1)}/\lambda_{1(2)})f \notin [\{F_{\nu_2,\nu_1}(\alpha/2)\}^{-1}, F_{\nu_1,\nu_2}(\alpha/2)]\big\} = 0.726 \\
&\text{and }\ \text{Power}(F_3) = P\big\{\gamma^{-1}h^{-1}(\lambda_{1(1)}/\lambda_{1(2)})f \notin [\{F_{\nu_2,\nu_1}(\alpha/2)\}^{-1}, F_{\nu_1,\nu_2}(\alpha/2)]\big\} \\
&\qquad\qquad\qquad = 0.908,
\end{aligned}$$

where f denotes a random variable distributed as F distribution with degrees of freedom, ν_1 and ν_2. Note that Power(F_2) and Power(F_3) give lower bounds of the asymptotic powers when $\tilde{h}_* = h^{-1}$ and $\tilde{\gamma}_* = \gamma^{-1}$.

In Figure 4, we summarized the findings obtained by averaging the outcomes from 4000 ($= R$, say) replications. Here, the first 2000 replications were generated by setting $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1$ as in (5.1) and the last 2000 replications were generated by setting $\mathbf{\Sigma}_2$ as in (5.2). Let F_{ir} $(i = 1, 2, 3)$ be the rth observation of F_i for $r = 1, ..., 4000$. We defined $P_r = 1$ (or 0) when H_0 was falsely rejected (or not) for $r = 1, ..., 2000$, and H_a was falsely rejected (or not) for $r = 2001, ..., 4000$. We defined $\overline{\alpha} = (R/2)^{-1} \sum_{r=1}^{R/2} P_r$ to estimate the size and



Sizes of F_1, F_2 and F_3 Powers of F_1, F_2 and F_3

Figure 4. The values of $\overline{\alpha}$ are denoted by the dashed lines in the left panel and the values of $1-\overline{\beta}$ are denoted by the dashed lines in the right panel for F_1, F_2 and F_3. The asymptotic powers were given by Power$(F_1)=0.39$, Power$(F_1)=0.726$ and Power$(F_3)=0.908$ which were denoted by the solid lines in the right panel.

$1-\overline{\beta}=1-(R/2)^{-1}\sum_{r=R/2+1}^{R}P_r$ to estimate the power. Their standard deviations are less than 0.011. Throughout, the tests gave adequate performances for the high-dimensional cases.

Appendix A.

Throughout, let $\boldsymbol{P}_n=\boldsymbol{I}_n-\boldsymbol{1}_n\boldsymbol{1}_n^T/n$, where $\boldsymbol{1}_n=(1,...,1)^T$. Let $\boldsymbol{e}_n=(e_1,...,e_n)^T$ be an arbitrary (random) n-vector such that $||\boldsymbol{e}_n||=1$ and $\boldsymbol{e}_n^T\boldsymbol{1}_n=0$.

Proof of Proposition 2.1. We assume $\boldsymbol{\mu}=\boldsymbol{0}$ without loss of generality. We write that $\boldsymbol{X}^T\boldsymbol{X}=\sum_{s=1}^{i_*}\lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T+\sum_{s=i_*+1}^{d}\lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T$ for $i_*=1$ when n is fixed, and for some fixed $i_*(\geq 1)$ when $n\to\infty$. Here, by using Markov's inequality, for any $\tau>0$, under (A-ii) and (A-iii), we have that

$$P\Big\{\sum_{j=1}^{n}\Big(\sum_{s=i_*+1}^{d}\frac{\lambda_s(z_{sj}^2-1)}{n\lambda_1}\Big)^2>\tau\Big\}\leq\frac{\sum_{r,s\geq 2}^{d}\lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\}}{\tau n\lambda_1^2}\to 0$$

and $$P\Big\{\sum_{j\neq j'}^{n}\Big(\sum_{s=i_*+1}^{d}\frac{\lambda_s z_{sj}z_{sj'}}{n\lambda_1}\Big)^2>\tau\Big\}\leq\frac{\delta_{i_*}}{\tau\lambda_1^2}\to 0$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Note that $\sum_{j=1}^n e_j^4 \le 1$ and $\sum_{j\neq j'}^n e_j^2 e_{j'}^2 \le 1$. Then, under (A-ii) and (A-iii), we have that

$$\begin{aligned}
\Big|\sum_{j=1}^n e_j^2 \sum_{s=i_*+1}^d \frac{\lambda_s(z_{sj}^2-1)}{n\lambda_1}\Big| &\le \Big\{\sum_{j=1}^n e_j^4\Big\}^{1/2}\Big\{\sum_{j=1}^n\Big(\sum_{s=i_*+1}^d \frac{\lambda_s(z_{sj}^2-1)}{n\lambda_1}\Big)^2\Big\}^{1/2} \\
&= o_p(1) \quad \text{and} \\
\Big|\sum_{j\neq j'}^n e_j e_{j'} \sum_{s=i_*+1}^d \frac{\lambda_s z_{sj} z_{sj'}}{n\lambda_1}\Big| &\le \Big\{\sum_{j\neq j'}^n e_j^2 e_{j'}^2\Big\}^{1/2}\Big\{\sum_{j\neq j'}^n\Big(\sum_{s=i_*+1}^d \frac{\lambda_s z_{sj} z_{sj'}}{n\lambda_1}\Big)^2\Big\}^{1/2} \\
&= o_p(1)
\end{aligned}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Thus, we claim that

$$\boldsymbol{e}_n^T \frac{\boldsymbol{X}^T\boldsymbol{X}}{(n-1)\lambda_1}\boldsymbol{e}_n = \boldsymbol{e}_n^T \frac{\sum_{s=1}^{i_*}\lambda_s \boldsymbol{z}_s \boldsymbol{z}_s^T}{(n-1)\lambda_1}\boldsymbol{e}_n + \frac{\kappa}{(n-1)\lambda_1} + o_p(1) \tag{A.1}$$

from the fact that $\sum_{s=i_*+1}^d \lambda_s/(n\lambda_1) = \kappa/(n\lambda_1) + o(1)$ when $n \to \infty$. Note that $\boldsymbol{e}_n^T\boldsymbol{P}_n = \boldsymbol{e}_n^T$ and $\boldsymbol{P}_n\boldsymbol{z}_s = \boldsymbol{z}_{os}$ for all s. Also, note that $(\boldsymbol{z}_{os}/n^{1/2})^T(\boldsymbol{z}_{os'}/n^{1/2}) = o_p(1)$ for $s \neq s'$ as $n \to \infty$ from the fact that $E\{(\boldsymbol{z}_{os}^T\boldsymbol{z}_{os'}/n)^2\} = o(1)$ as $n \to \infty$. Then, by noting that $P(\lim_{d\to\infty}||\boldsymbol{z}_{o1}|| \neq 0) = 1$, $\liminf_{d\to\infty}\lambda_1/\lambda_2 > 1$ and $\boldsymbol{z}_{o1}^T\boldsymbol{1}_n = 0$, it holds that

$$\begin{aligned}
\max_{\boldsymbol{e}_n}\Big\{\boldsymbol{e}_n^T \frac{\sum_{s=1}^{i_*}\lambda_s \boldsymbol{z}_s\boldsymbol{z}_s^T}{(n-1)\lambda_1}\boldsymbol{e}_n\Big\} &= \max_{\boldsymbol{e}_n}\Big\{\boldsymbol{e}_n^T \frac{\sum_{s=1}^{i_*}\lambda_s \boldsymbol{z}_{os}\boldsymbol{z}_{os}^T}{(n-1)\lambda_1}\boldsymbol{e}_n\Big\} \\
&= ||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 + o_p(1)
\end{aligned} \tag{A.2}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Note that $\hat{\boldsymbol{u}}_1^T\boldsymbol{1}_n = 0$ and $\hat{\boldsymbol{u}}_1^T\boldsymbol{P}_n = \hat{\boldsymbol{u}}_1^T$ when $\boldsymbol{S}_D \neq \boldsymbol{O}$. Then, from (A.1), (A.2) and $\boldsymbol{P}_n\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{P}_n/(n-1) = \boldsymbol{S}_D$, under (A-ii) and (A-iii), we have that

$$\hat{\boldsymbol{u}}_1^T\frac{\boldsymbol{S}_D}{\lambda_1}\hat{\boldsymbol{u}}_1 = \hat{\boldsymbol{u}}_1^T\frac{\boldsymbol{X}^T\boldsymbol{X}}{(n-1)\lambda_1}\hat{\boldsymbol{u}}_1 = ||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 + \frac{\kappa}{(n-1)\lambda_1} + o_p(1) \tag{A.3}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. It concludes the result. $\square$

Proof of Lemma 2.1. By using Markov's inequality, for any $\tau > 0$, under (A-ii)

and (A-iii), we have that

$$
\begin{aligned}
&P\Big\{\Big(\sum_{s=2}^{d}\frac{\lambda_s\{||\boldsymbol{z}_{os}||^2-(n-1)\}}{(n-1)\lambda_1}\Big)^2>\tau\Big\}\\
&=P\Big\{\Big(\sum_{s=2}^{d}\frac{\lambda_s\{(n-1)\sum_{k=1}^{n}(z_{sk}^2-1)/n-\sum_{k\neq k'}^{n}z_{sk}z_{sk'}/n\}}{(n-1)\lambda_1}\Big)^2>\tau\Big\}\\
&=O\Big\{\frac{\sum_{r,s\geq 2}^{d}\lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\}}{n\lambda_1^2}\Big\}+O\{\delta_1/(n\lambda_1)^2\}\to 0
\end{aligned}
$$

as $d\to\infty$ either when n is fixed or $n\to\infty$. Thus it holds that tr$(\boldsymbol{S}_D)/\lambda_1=\kappa/\lambda_1+||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2+o_p(1)$ from the fact that tr$(\boldsymbol{S}_D)=\lambda_1||\boldsymbol{z}_{o1}||^2/(n-1)+\sum_{s=2}^{d}\lambda_s||\boldsymbol{z}_{os}||^2/(n-1)$. Then, from Proposition 2.1 and $\liminf_{d\to\infty}\kappa/\lambda_1>0$, we can claim the results. □

Proof of Theorem 2.1. When $n\to\infty$, we can claim the results from Theorems 4.1, 4.2 and Corollary 4.1 in Yata and Aoshima (2013). When n is fixed, by combining Proposition 2.1 with Lemma 2.1, we can claim the results because $||\boldsymbol{z}_{o1}||^2=\sum_{k=1}^{n}z_{1k}^2-n\bar{z}_1^2$ is distributed as χ_{n-1}^2 if $z_{1j},\ j=1,...,k,$ are i.i.d. as $N(0,1)$. □

Proof of Theorem 2.2. From Theorem 2.1 and Lemma 2.1, under (A-i) to (A-iii), it holds that

$$
\begin{aligned}
&P\Big(\frac{\lambda_1}{\mathrm{tr}(\boldsymbol{\Sigma})}\in\Big[\frac{(n-1)\tilde{\lambda}_1}{b\tilde{\kappa}+(n-1)\tilde{\lambda}_1},\frac{(n-1)\tilde{\lambda}_1}{a\tilde{\kappa}+(n-1)\tilde{\lambda}_1}\Big]\Big)\\
&=P\Big(\frac{(n-1)\tilde{\lambda}_1}{b\tilde{\kappa}+(n-1)\tilde{\lambda}_1}\leq\frac{\lambda_1}{\mathrm{tr}(\boldsymbol{\Sigma})}\leq\frac{(n-1)\tilde{\lambda}_1}{a\tilde{\kappa}+(n-1)\tilde{\lambda}_1}\Big)\\
&=P\Big(\frac{a\tilde{\kappa}}{(n-1)\tilde{\lambda}_1}\leq\frac{\kappa}{\lambda_1}\leq\frac{b\tilde{\kappa}}{(n-1)\tilde{\lambda}_1}\Big)=P\Big(a\leq(n-1)\frac{\tilde{\lambda}_1\kappa}{\lambda_1\tilde{\kappa}}\leq b\Big)\\
&=1-\alpha+o(1)
\end{aligned}
$$

as $d\to\infty$ when n is fixed. It concludes the result. □

Proofs of Lemmas 3.1 and 3.2. We note that $||\boldsymbol{z}_{o1}||^2/n=1+o_p(1)$ as $n\to\infty$. From (A.3), under (A-ii) and (A-iii), we have that

$$\hat{\boldsymbol{u}}_1^T\boldsymbol{z}_{o1}/||\boldsymbol{z}_{o1}||=1+o_p(1)\qquad\text{(A.4)}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$, so that $\hat{\boldsymbol{u}}_1^T \boldsymbol{z}_{o1} = ||\boldsymbol{z}_{o1}|| + o_p(n^{1/2})$. Thus, we can claim the result of Lemma 3.2. On the other hand, with the help of Proposition 2.1, under (A-ii) and (A-iii), it holds that from (A.4)

$$\boldsymbol{h}_1^T \hat{\boldsymbol{h}}_1 = \frac{\boldsymbol{h}_1^T(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_1}{\{(n-1)\hat{\lambda}_1\}^{1/2}} = \frac{\lambda_1^{1/2} \boldsymbol{z}_{o1}^T \hat{\boldsymbol{u}}_1}{\{(n-1)\hat{\lambda}_1\}^{1/2}} = \frac{||\boldsymbol{z}_{o1}|| + o_p(n^{1/2})}{\{||\boldsymbol{z}_{o1}||^2 + \kappa/\lambda_1 + o_p(n)\}^{1/2}}$$
$$= \frac{1}{\{1 + \kappa/(\lambda_1 ||\boldsymbol{z}_{o1}||^2)\}^{1/2}} + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. It concludes the result of Lemma 3.1. □

Proof of Theorem 3.1. With the help of Theorem 2.1, under (A-ii) and (A-iii), we have that from (A.4)

$$\boldsymbol{h}_1^T \tilde{\boldsymbol{h}}_1 = \frac{\boldsymbol{h}_1^T(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_1}{\{(n-1)\tilde{\lambda}_1\}^{1/2}} = \frac{||\boldsymbol{z}_{o1}|| + o_p(n^{1/2})}{\{||\boldsymbol{z}_{o1}||^2 + o_p(n)\}^{1/2}} = 1 + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. It concludes the result. □

Proof of Theorem 3.2. By combing Theorem 2.1 with Lemma 3.2, under (A-ii) and (A-iii), we have that

$$\tilde{s}_{1j}/\sqrt{\lambda_1} = \hat{u}_{1j}\sqrt{(n-1)\tilde{\lambda}_1/\lambda_1} = \hat{u}_{1j}||\boldsymbol{z}_{o1}|| + o_p(1) = z_{o1j} + o_p(1)$$

as $d \to \infty$ when n is fixed. By noting that $z_{o1j} = z_{1j} - \bar{z}_1$ and $\bar{z}_1$ is distributed as $N(0, 1/n)$ under (A-i), we have the results. □

Proof of Corollary 4.1. From Theorem 2.1, the result is obtained straightforwardly. □

Proof of Lemma 4.1. Let $\boldsymbol{Z}_i = [\boldsymbol{z}_{1(i)}, ..., \boldsymbol{z}_{d(i)}]^T$ be a sphered data matrix of π_i for $i = 1, 2$, where $\boldsymbol{z}_{j(i)} = (z_{j1(i)}, ..., z_{jn_i(i)})^T$. We assume $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{0}$ without loss of generality. Let $\beta_{st} = (\lambda_{s(1)}\lambda_{t(2)})^{1/2}\boldsymbol{h}_{s(1)}^T\boldsymbol{h}_{t(2)}$ for all s, t. Let $i_\star$ be a fixed constant such that $\sum_{s=i_\star+1}^{d} \lambda_{s(j)}^2/\lambda_{1(j)}^2 = o(1)$ as $d \to \infty$ for $j = 1, 2$. Note that

$i_\star$ exists under (A-ii) for each π_i. We write that

$$\boldsymbol{X}_1^T\boldsymbol{X}_2 = \sum_{s,t\le i_\star}\beta_{st}\boldsymbol{z}_{s(1)}\boldsymbol{z}_{t(2)}^T + \sum_{s,t\ge i_\star+1}^{d}\beta_{st}\boldsymbol{z}_{s(1)}\boldsymbol{z}_{t(2)}^T$$
$$+ \sum_{s=i_\star+1}^{d}\sum_{t=1}^{i_\star}\beta_{st}\boldsymbol{z}_{s(1)}\boldsymbol{z}_{t(2)}^T + \sum_{s=1}^{i_\star}\sum_{t=i_\star+1}^{d}\beta_{st}\boldsymbol{z}_{s(1)}\boldsymbol{z}_{t(2)}^T.$$

Note that

$$E\Big\{\Big(\sum_{s=i_\star+1}^{d}\sum_{t=1}^{i_\star}\beta_{st}z_{sj(1)}z_{tj'(2)}\Big)^2\Big\}$$
$$= \mathrm{tr}\Big(\sum_{s=i_\star+1}^{d}\lambda_{s(1)}\boldsymbol{h}_{s(1)}\boldsymbol{h}_{s(1)}^T\sum_{t=1}^{i_\star}\lambda_{t(2)}\boldsymbol{h}_{t(2)}\boldsymbol{h}_{t(2)}^T\Big) \le i_\star\lambda_{i_\star+1(1)}\lambda_{1(2)}$$

for all j,j'. Also, note that

$$E\Big\{\Big(\sum_{s,t\ge i_\star+1}^{d}\beta_{st}z_{sj(1)}z_{tj'(2)}\Big)^2\Big\} = \mathrm{tr}\Big(\sum_{s=i_\star+1}^{d}\lambda_{s(1)}\boldsymbol{h}_{s(1)}\boldsymbol{h}_{s(1)}^T\sum_{t=i_\star+1}^{d}\lambda_{t(2)}\boldsymbol{h}_{t(2)}\boldsymbol{h}_{t(2)}^T\Big)$$
$$\le \Big(\sum_{s=i_\star+1}^{d}\lambda_{s(1)}^2\sum_{t=i_\star+1}^{d}\lambda_{t(2)}^2\Big)^{1/2}$$

for all j,j'. Then, by using Markov's inequality, for any $\tau>0$, under (A-ii) for each π_i, we have that

$$P\Big\{\sum_{j=1}^{n_1}\sum_{j'=1}^{n_2}\Big(\sum_{s=i_\star+1}^{d}\sum_{t=1}^{i_\star}\frac{\beta_{st}z_{sj(1)}z_{tj'(2)}}{(n_1n_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}}\Big)^2>\tau\Big\}\to 0,$$
$$P\Big\{\sum_{j=1}^{n_1}\sum_{j'=1}^{n_2}\Big(\sum_{s=1}^{i_\star}\sum_{t=i_\star+1}^{d}\frac{\beta_{st}z_{sj(1)}z_{tj'(2)}}{(n_1n_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}}\Big)^2>\tau\Big\}\to 0$$
$$\text{and } P\Big\{\sum_{j=1}^{n_1}\sum_{j'=1}^{n_2}\Big(\sum_{s,t\ge i_\star+1}^{d}\frac{\beta_{st}z_{sj(1)}z_{tj'(2)}}{(n_1n_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}}\Big)^2>\tau\Big\}\to 0$$

as $d\to\infty$ either when n_i is fixed or $n_i\to\infty$ for $i=1,2$. Hence, similar to (A.1), it holds that

$$\frac{\boldsymbol{e}_{n_1}^T\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{e}_{n_2}}{(\nu_1\nu_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}} = \frac{\boldsymbol{e}_{n_1}^T\sum_{s,t\le i_\star}\beta_{st}\boldsymbol{z}_{s(1)}\boldsymbol{z}_{t(2)}^T\boldsymbol{e}_{n_2}}{(\nu_1\nu_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}} + o_p(1).$$

Note that $\boldsymbol{e}_{n_i}^T \boldsymbol{P}_{n_i} = \boldsymbol{e}_{n_i}^T$ and $\boldsymbol{P}_{n_i} \boldsymbol{z}_{1(i)} = \boldsymbol{z}_{o1(i)}$ for $i = 1, 2$, where $\boldsymbol{z}_{o1(i)} = \boldsymbol{z}_{1(i)} - (\bar{z}_{1(i)}, ..., \bar{z}_{1(i)})^T$ and $\bar{z}_{1(i)} = n_i^{-1} \sum_{k=1}^{n_i} z_{1k(i)}$. Also, note that $\boldsymbol{X}_i \boldsymbol{P}_{n_i} = (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)$ for $i = 1, 2$, where $\overline{\boldsymbol{X}}_i = [\bar{\boldsymbol{x}}_i, ..., \bar{\boldsymbol{x}}_i]$ and $\bar{\boldsymbol{x}}_i = \sum_{j=1}^{n_i} \boldsymbol{x}_{j(i)}/n_i$. Let $\hat{\boldsymbol{u}}_{1(i)}$ be the first (unit) eigenvector of $(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)^T(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)$ for $i = 1, 2$. Note that $\hat{\boldsymbol{u}}_{1(i)}^T \boldsymbol{P}_{n_i} = \hat{\boldsymbol{u}}_{1(i)}^T$ when $(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)^T(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i) \neq \boldsymbol{O}$ for $i = 1, 2$. Then, under (A-ii) for each π_i, we have that

$$\frac{\hat{\boldsymbol{u}}_{1(1)}^T(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2)\hat{\boldsymbol{u}}_{1(2)}}{(\nu_1\nu_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}} = \frac{\hat{\boldsymbol{u}}_{1(1)}^T \sum_{s,t \le i_\star} \beta_{st} \boldsymbol{z}_{os(1)} \boldsymbol{z}_{ot(2)}^T \hat{\boldsymbol{u}}_{1(2)}}{(\nu_1\nu_2\lambda_{1(1)}\lambda_{1(2)})^{1/2}} + o_p(1) \tag{A.5}$$

as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$. Note that $\tilde{\boldsymbol{h}}_{1(i)} = \{\nu_i \tilde{\lambda}_{1(i)}\}^{-1/2}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)\hat{\boldsymbol{u}}_{1(i)}$ for $i = 1, 2$. Also, note that $\boldsymbol{z}_{os(i)}^T \boldsymbol{z}_{os'(i)}/n_i = o_p(1)$ $(s \neq s')$ when $n_i \to \infty$ for $i = 1, 2$. Then, by combining (A.5) with Theorem 2.1 and (A.4), we can claim the result. □

Proofs of Theorems 4.1 and 4.2. By combining Theorem 2.1, Lemmas 2.1 and 4.1, we can claim the results. □

Acknowledgements

Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 26800078. Research of the third author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 26540010.

References

Ahn, J., Marron, J.S., Muller, K.M., Chi, Y.-Y., 2007. The high-dimension, low-sample-size geometric representation holds under mild conditions. Biometrika 94, 760-766.

Aoshima, M., Yata, K., 2011. Two-stage procedures for high-dimensional data. Sequential Anal. (Editor's special invited paper) 30, 356-399.

Aoshima, M., Yata, K., 2013. Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. Methodol. Comput. Appl. Probab., in press (DOI: 10.1007/s11009-013-9370-7).

Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R., Korsmeyer, S.J., 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nature Genetics 30, 41-47.

Hall, P., Marron, J.S., Neeman, A., 2005. Geometric representation of high dimension, low sample size data. J. R. Statist. Soc. B 67, 427-444.

Ishii, A., Yata, K., Aoshima, M., 2014. Asymptotic distribution of the largest eigenvalue via geometric representations of high-dimension, low-sample-size data. Sri Lankan J. Appl. Statist., Special Issue: Modern Statistical Methodologies in the Cutting Edge of Science (ed. Mukhopadhyay, N.), 81-94.

Jung, S., Marron, J.S., 2009. PCA consistency in high dimension, low sample size context. Ann. Statist. 37, 4104-4130.

Jung, S., Sen, A., Marron, J.S., 2012. Boundary behavior in high dimension, low sample size asymptotics of PCA. J. Multivariate Anal. 109, 190-203.

Srivastava, M.S., Yanagihara, H., 2010. Testing the equality of several covariance matrices with fewer observations than the dimension. J. Multivariate Anal. 101, 1319-1329.

Yata, K., Aoshima, M., 2009. PCA consistency for non-Gaussian data in high dimension, low sample size context. Commun. Statist. Theory Methods, Special Issue Honoring Zacks, S. (ed. Mukhopadhyay, N.) 38, 2634-2652.

Yata, K., Aoshima, M., 2010. Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. J. Multivariate Anal. 101, 2060-2077.

Yata, K., Aoshima, M., 2012. Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. J. Multivariate Anal. 105, 193-215.

Yata, K., Aoshima, M., 2013. PCA consistency for the power spiked model in high-dimensional settings. J. Multivariate Anal. 122, 334-354.